

UDK 004.4

Kulykovska N.A.¹, Ostakh P.O.²

¹ senior teacher NU «Zaporizhzhia Polytechnic»

² student of group CST -622m NU «Zaporizhzhia Polytechnic»

COMPARISON OF BIG DATA TECHNOLOGIES

The popularity of big data technologies can be attributed to several reasons, including the explosion of data, the need for competitive advantage, and the potential for innovation. As a result, many companies have developed various technologies for handling big data, including Hadoop, Spark, NoSQL databases, and data warehousing solutions such as Amazon Redshift and Google BigQuery. Apache Software Foundation, Cloudera, Hortonworks, IBM, Google, Amazon Web Services, and Oracle are among the leading companies developing big data technologies.

With the growth of the internet and the proliferation of smart devices, the amount of data being generated has exploded. Big data technologies help organizations to process and analyze this data to gain insights and make informed decisions. Big data technologies enable faster processing of large amounts of data, allowing organizations to analyze data in real-time and respond quickly to changes in their business. Big data technologies enable the processing of complex data types, such as unstructured data, which was previously difficult or impossible to analyze. This allows organizations to gain a more comprehensive understanding of their data and make more accurate decisions. With the emergence of cloud computing and open-source technologies, big data technologies have become more accessible and cost-effective for organizations of all sizes.

There are many software options for working with big data. Hadoop is a distributed computing framework designed to handle large data sets, while Spark is an open-source data processing engine that provides fast and flexible data processing capabilities. NoSQL databases like Cassandra and MongoDB are

designed to handle large amounts of unstructured data. Amazon Redshift and Google BigQuery provide cloud-based data warehousing solutions that can handle large data sets. Apache Flink and Apache Storm are other open-source data processing frameworks that can be used for real-time data processing.

When it comes to programming languages, Java is widely used in the development of big data technologies like Hadoop and Spark due to its performance and scalability. Python is also a popular choice for big data analytics due to its ease of use and support for a wide range of data processing libraries. R is a statistical programming language that is commonly used for data analysis and machine learning (tabl.1).

The potential benefits of big data technologies make them an increasingly popular choice for businesses and organizations looking to gain a competitive edge and drive innovation.

Table 1 – Comparison of big data technologies

<i>Main technologies:</i>	
Hadoop	Developed by Apache Software Foundation, Hadoop is a distributed computing framework designed to handle large data sets.
Spark	Developed by Apache Software Foundation, Spark is an open-source data processing engine that provides fast and flexible data processing capabilities.
NoSQL	These databases are designed to handle large amounts of unstructured data and can be used in conjunction with big data processing frameworks like Hadoop and Spark.
Data warehousing	Technologies like Amazon Redshift and Google BigQuery provide cloud-based data warehousing solutions that can handle large data sets.
<i>Programming languages</i>	
Java	Java is widely used in the development of big data technologies like Hadoop and Spark.
Python	Python is a popular choice for big data analytics due to its ease of use and support for a wide range of data processing libraries.
R	R is a statistical programming language that is commonly used for data analysis and machine learning.
<i>Advantages of big data</i>	
Scalability	Big data technologies are designed to handle large amounts of data and can scale to meet the needs of businesses of all sizes.
Flexibility	Big data technologies provide a flexible and agile approach to data processing and analysis, allowing businesses to quickly adapt to changing business needs.
Improved decision-making	Big data technologies can provide businesses with insights into customer behavior, market trends.

<i>Disadvantages of big data</i>	
Complexity	Big data technologies can be complex to set up and manage, requiring specialized expertise and resources.
Security	Managing and securing large data sets can be challenging, and there are risks associated with data breaches and other security threats.
Cost	The cost of implementing and maintaining big data technologies can be high, particularly for small businesses with limited resources.